HUAWEI TECHNOLOGIES CO., LTD.
Huawei Industrial Base
Bantian Longgang
Shenzhen 518129, P. R. China
Tel: +86-755-28780808
www.huawei.com

HUAWEI

Striding Towards the Intelligent World White Paper 2024

Cloud Core Network

Intelligence + 5G-A: Elevating Connectivity Beyond Boundaries

Building a Fully Connected, Intelligent World

Cloud Core Network

Abstract

The year 2024 marks a significant milestone in the evolution of mobile networks, as it introduces the first year of commercial use of 5G-A and the integration of AI with devices. This symbolizes the dawn of the mobile AI era, where intelligent services will become ubiquitous. With the completion of 3GPP Release 18, which defined 5G-A for the first time, more than 60 global operators and industry partners have reached a consensus on its value and begun commercialization. Leading operators are now actively exploring ways to leverage the new capabilities of AI and 5G-A to create new business value.

Building on the influential white paper "Striding Towards the Intelligent World 2023" by Huawei Cloud Core Network, which outlined seven key trends in core network evolution, this white paper presents a comprehensive analysis of the latest innovations, practices, and insights in the 5G-A core network. Through collaboration with industry partners, we have identified new trends and opportunities for operators to leverage cutting-edge AI and 5G-A capabilities to unlock new business value. This white paper presents six new trends of core network evolution from three complementary perspectives: the fusion of AI with devices, networks, and services; the new network capabilities enabled by 5G-A; and the evolution of the telco cloud. Our goal is to provide a valuable reference for the industry, sharing our thoughts and suggestions on how to drive development and innovation in the years to come.

2023

Trend 1: MEC brings extensive B2B opportunities to operators.	New Connectivity
Trend 2: New technologies such as RedCap, A-IoT, and ISAC have entered the standardization phase, which heralds the incubation of relevant scenario-based applications.	
Trend 3: Large-scale commercial use of 5G accelerates VoLTE monetization, and VoLTE is on the road to experience upgrade.	New Calling
Trend 4: Communications capabilities further opened up; New Calling becomes typical application.	

Trend 5: Glasses-free 3D is boosting mobile media traffic; ultra-large traffic processing and intelligent experience assurance are becoming trending topics.	New Video
Trend 6: 5G core networks are 100% deployed over Telco Cloud, cloudification becomes the leading network construction model, and future-oriented convergence & simplification, standardization, and automation become the directions.	Telco Cloud
Trend 7: Foundation models are turbocharging AI progress, indispensable for 5.5G Core to achieve intelligent O&M, services, and networks.	Intelligence

2024

Trend 1: 5G-A satisfies diversified service requirements and enables experience-based operations with ultra-high bandwidth and ultra-low latency.	Network Intelligence (Experience-based operations)
Trend 2: New Calling enables operators to revitalize calling services and capture dialer-based business opportunities in the mobile AI era.	Service Intelligence (New Calling)
Trend 3: Operators need to leverage AI to enhance O&M efficiency, boost network reliability, and address "historical debts."	O&M Intelligence (Digital employees)
Trend 4: 5G-A new capabilities offer higher bandwidth, deeper industry integration, meshed networks, and more connections to deliver intelligent, deterministic connections.	Enhanced Connectivity
Trend 5: 5G-A goes beyond connections; it integrates communications, computing, and storage, facilitating content storage on operator networks and extending into low-altitude economy.	Beyond Connectivity
Trend 6: A decade of NFV exploration leads operators to embrace telco cloud as a strategic imperative to realize E2E automation.	Telco Cloud

Insights into the Core Network Trends of 2024

Trend 1

5G-A Satisfies Diversified Service Requirements and Enables Experience-based Operations with Ultra-High Bandwidth and Ultra-Low Latency

Since the commercial launch of 5G five years ago, the technology has generated significant business value and social benefits. Pioneer operators initially capitalized on the first wave of dividends through traffic-based service monetization. However, the industry now faces new challenges: slowing mobile traffic growth and stagnant revenue. As a result, operators are under pressure to develop a new growth curve and transform their business models. This shift from traffic-based to experience-based monetization has become inevitable, requiring operators to rethink their strategies and prioritize delivering exceptional user experiences to stay competitive.

The commercialization of 5G-A and the mobile AI revolution have given rise to a new generation of services, demanding improved experiences and differentiated assurance. This establishes a solid service foundation for experience-based operations. The 5G-A network, with its distinctive features of super uplink and downlink bandwidths and low latencies, provides a robust network foundation for delivering these experiences. Furthermore, market research indicates that a significant majority (over 55%) of global consumers are willing to pay more for differentiated service experiences, creating a strong user foundation for experience-based operations.

Trend 2

New Calling Enables Operators to Revitalize Calling Services and Capture Dialer-based Business Opportunities in the Mobile AI Era

In 2023, leading operators initiated a strategic shift towards service intelligence with New Calling, aiming to revolutionize calling services, elevate user experiences, strengthen their user base, and ultimately drive revenue growth. In the following year, the rise of 5G-A

and its integration with AI sparked a transformation across the calling industry, shifting from transparent information exchange to multi-modal communications and presenting significant opportunities for operators.

In the second half of 2023, China Mobile launched a New Calling network across China and rapidly acquired tens of millions of subscribers within just six months. Meanwhile, operators across Europe, Latin America, the Middle East, and Asia Pacific have piloted New Calling services, establishing a strong foundation for global expansion.

In the competitive calling market, leading operators are prioritizing the integration of AI with calling services. By deploying AI-driven intelligent calling assistants, they enhance communication experiences and efficiency. This strategic approach enables third-party AI foundation models to accelerate user acquisition, and operators can build user-friendly, dialer-based service gateways. As a flagship New Calling service, visualized voice calling revolutionizes mobile calling by integrating AIGC and rich media capabilities, empowering operators to transition from voice-centric to content-driven business models and unlock new revenue streams.

In June 2024, 3GPP Release 18 was officially finalized, marking a significant milestone in the evolution of calling services. Release 18 defines the architecture, interfaces, and service procedures of New Calling networks, providing a crucial technical foundation for the industry's growth. At the end of 2023, GSMA's Board of Directors established the New Calling Task Force, with a mission to promote the construction of a New Calling standards ecosystem. Thanks to the unwavering dedication of the Task Force, DC-native terminals are poised to enter the market by the end of 2024, setting the stage for a rapid and widespread adoption of interactive calling services.

Trend 3

Operators Need to Leverage AI to Enhance O&M Efficiency, Boost Network Reliability, and Address "Historical Debts"

The continuous evolution of Cloud Native and the introduction of new RATs like 5G have significantly increased O&M complexity and workload. Despite this, the number of O&M personnel has remained static, creating a widening gap between the workload and available staff. Consequently, some non-critical alarms are being ignored, posing network reliability risks. Furthermore, the accumulation of "historical debts" is a pressing concern that demands immediate attention. However, the growth of foundation models offers new hope, presenting potential solutions to traditional O&M challenges.

In response to these challenges, operators urgently need to leverage AI and deploy "digital employees" that can enhance network O&M efficiency and address "historical debts." Against this backdrop, there is a growing focus on creating a multi-modal O&M foundation model for the core network. This involves building role-based copilots and scenario-oriented agents, as well as coordinating foundation and small models. The ultimate goal is to accelerate progress towards L4 autonomous networks.

Trend 4

5G-A New Capabilities Offer Higher Bandwidth, Deeper Industry Integration, Meshed Networks, and More Connections to Deliver Intelligent, Deterministic Connections

Higher bandwidth: The rise of HD and intelligent mobile Internet services has significantly increased the demand for ultra-high bandwidth. For instance, cloud gaming, XR applications, 4K live streaming, and speed test software require minimum bandwidths of 100 Mbit/s, 200 Mbit/s, 800 Mbit/s, and 1 Gbit/s, respectively. Also, the development of 3CC technology for base stations and terminals has overcome the bottleneck in air interface transmission. Consequently, the core network needs to deliver higher forwarding rates to satisfy E2E high bandwidth requirements.

Deeper industry integration: GSMA forecasts that the profits of 5G industry private networks will jump to US$109 billion in 2030, with a staggering compound annual growth rate of 66% over five years. Leveraging PNI-NPN has become a consensus among operators. To date, over 15,000 5G private networks have been deployed worldwide, covering sectors such as education, healthcare, government, and energy. 5G is evolving from a peripheral tool to a core enabler for key production processes across various industries.

Meshed networks: As service demands escalate, networks are becoming increasingly multidimensional. The advancement of intelligent terminals and networks in turn enable users to access network services through different terminals anytime, anywhere. For example, users need to remotely access smart home or enterprise office devices, access enterprise private networks, or remotely access a third vehicle-centric space. This trend is propelling networks to evolve from B2C-only connections to fully meshed connections among humans, vehicles, homes, and enterprises.

More connections: Diverse IoT service scenarios have different requirements for IoT networks. In the 4G era, NB-IoT and LTE-M have emerged to meet the demands of the tens of billions of medium-speed IoT applications. However, in the 5G/5G-A era, RedCap unlocks high-speed IoT access, while A-IoT pioneers ultra-low-speed IoT applications, paving the way for 100 billion IoT connections.

Trend 5

5G-A Goes Beyond Connections; It Integrates Communications, Computing, and Storage, Facilitating Content Storage on Operator Networks and Extending into Low-Altitude Economy

Short videos have become a dominant service on the mobile Internet. The global daily active user base has exceeded 2.6 billion in 2023, and Chinese users account for 90% of this total. As 5G rapidly expands, short videos are more likely to be available in HD, thanks to the prevalence of H.265 encoded videos, which constitute approximately 90% of all short video content. However, this shift to HD content significantly increases the bandwidth demand of OTT platforms. For example, to serve 3 million users, Douyin (the Chinese equivalent of TikTok) requires a CDN bandwidth of around 60 Gbit/s. Consequently, OTT platforms and telecom operators face escalating costs. To mitigate these expenses, OTT platforms may choose to lower the video bitrates, but this decision impacts user experience. Lower bitrate videos also result in reduced user data consumption, affecting both operators' revenue and the sales of premium data packages.

Content storage on operator networks offers a solution for both OTT platforms and operators. To achieve this, operators should delve deeper into media relay technologies. By optimizing video transmission efficiency, operators can alleviate CDN bandwidth pressure, enhance user experiences, boost data traffic, and increase operator revenue. Ultimately, this approach benefits users, OTT platforms, and operators alike.

HCS technologies enable the convergence of diverse capabilities onto a single network. With its multifaceted capabilities, it is being deployed in various scenarios such as low-altitude, maritime, and road transportation, serving as a significant growth engine for operators. The *China Low-Altitude Economy Development Report*, released by CCID Consulting, highlights the immense three-tier trillion CNY development potential of the low-altitude economy, encompassing low-altitude aircraft manufacturing, general airport construction, and low-altitude new infrastructure. From policy formulation, technological advancement, protocol research, to application in various scenarios, all facets continue to thrive in the HCS industry.

Trend 6

A Decade of NFV Exploration Leads Operators to Embrace Telco Cloud as a Strategic Imperative to Realize E2E Automation

Over the past decade, NFV technology, pioneered by ETSI, has made significant strides, ushering in the era of VNFs. As 5G-A and AI technologies continue to evolve, telecom networks are encountering new opportunities and challenges. These include the transition from VNFs to CNFs, the shift from general-purpose computing power (CPUs) to diversified computing power (XPUs), and the integration of AI. In response, operators, adhering to the ETSI standard framework, best IT practices, and carrier-grade requirements, are committed to building a robust telco infrastructure while consistently promoting industry innovation and development.

Contents

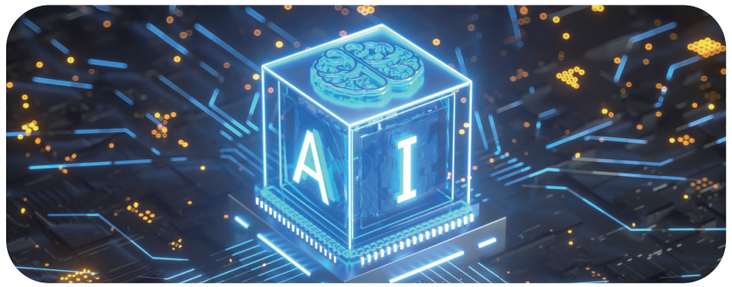
1

Overview

As 5G-A enters its inaugural year of commercial deployment, it converges with the emergence of on-device AI, propelling both technologies to unprecedented heights. This synergy gives rise to the mobile AI era, where intelligent services abound. With the finalization of 3GPP Release 18, which formally defines 5G-A, the industry stands at a critical juncture in adopting this transformative technology. Over 60 operators and industry partners have united to lay the foundation for 5G-A's commercialization, recognizing its immense potential. As the industry embarks on this new journey, harnessing the cutting-edge capabilities of AI and 5G-A becomes essential for navigating the disruptive changes and creating new business value.

The advent of on-device AI has revolutionized the mobile device landscape, fundamentally transforming the way users interact with their devices. Gone are the days of simple touch-based control; today, devices are capable of detecting and responding to a wide range of multi-modal cues, including tactile, auditory, and visual inputs, as well as natural language, gestures, and facial expressions. This paradigm shift has redefined the role of mobile terminals, elevating them from mere smartphones to AI-fortified personal assistants and embodied agents that seamlessly integrate into a range of diverse scenarios.

The integration of AI into networks has profound implications for business models and revenue streams. As traditional traffic-based operations give way to experience-oriented operations, operators must adapt to meet evolving user needs and deliver tailored



experiences across disparate scenarios. This shift requires a fundamental transformation of business models, as operators transition from relying solely on traffic volume to generating revenue through the delivery of high-quality, user-centric experiences.

The intersection of AI and services has reshaped digital service entry points, characterized by AI-driven interfaces that rival app-centric models. No longer must users download apps from the app store to access related services; instead, AI-powered services are accessible directly on terminals. A compelling example of AIGC-enabled evolution is the transformation of calling services, which have penetrated high-value B2C and B2B scenarios. Enterprises, in particular, are upgrading traditional customer service hotlines to AI-powered assistants, redefining the customer experience and driving new levels of engagement and loyalty.

The mobile AI era has sparked off a transformative convergence of AI and the 5G-A core network, revolutionizing connectivity and creating value beyond its traditional boundaries. As operators pursue a network transformation strategy focused on revenue growth, cost reduction, and user experience improvement, the fusion of AI with devices, networks, and services is driving unprecedented growth and innovation.

The 5G-A core network is characterized by its intelligence, encompassing three key areas: service intelligence, network intelligence, and O&M intelligence. Service intelligence enables operators to redefine service entry points, network intelligence enables experience-based operations and the reshaping of business models, and O&M intelligence reconstructs the O&M framework. With digital employees and O&M experts embedded, the new O&M framework helps achieve high network stability and efficiency.

The 5G-A core network is poised to deliver enhanced connectivity, featuring higher bandwidth, deeper industry integration, meshed networks, and more connections. Furthermore, by integrating cutting-edge HCS, computing, and storage capabilities with the 5G-A core network, operators can explore novel scenarios, especially those related to the low-altitude economy. This integration unlocks new business value in the mobile AI era.

By harnessing the power of the 5G-A core network, operators will be able to tap into a wide range of new business opportunities, from enhanced customer experiences to increased revenue streams. This white paper will delve deeper into the implications of this convergence and explore the strategies and solutions required to leverage the full potential of the 5G-A core network in the mobile AI era.

2

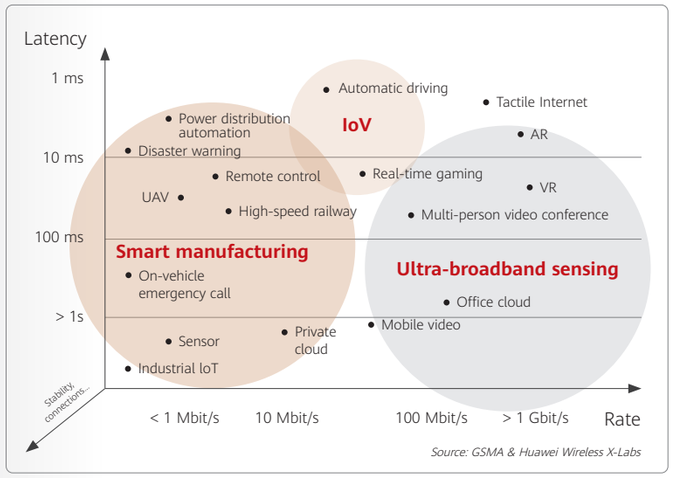
Trend 1

5G-A Satisfies Diversified Service Requirements and Enables Experience-based Operations with Ultra-High Bandwidth and Ultra-Low Latency

2.1 Service Foundation for Experience-based Operations: Emerging Services Like Immersive Applications, Ubiquitous AI, and Autonomous Driving Demand Lower Latency and Higher Bandwidth

The advent of 5G networks is driving a surge in innovative services such as XR, cloud gaming, cloud phones, and autonomous vehicles. Each of these services imposes unique demands on network bandwidth and latency, presenting new challenges for operators to deliver high-quality experiences.

Cloud gaming, for instance, requires a network latency of less than 20 milliseconds to ensure immediate operations and feedback, while XR applications demand an even lower latency of less than 10 milliseconds to prevent user dizziness and discomfort. Apollo Go, China's pioneering ride-hailing service, exemplifies the use of real-time data from screens to remotely control self-driving vehicles with precision. To guarantee the safety of both the vehicles and passengers, particularly in high-pressure situations, the platform must respond swiftly and accurately, necessitating a one-way latency of 5-20 milliseconds and an uplink bandwidth of up to 100 Mbit/s.



Ultra-broadband sensing

VR, AR, video conference, mobile video, etc.

IoV

Automatic driving, remote driving, fleet cruising, etc.

Smart manufacturing

Remote control, high-speed railway, power distribution, automation, sensor, etc.

New scenario exploration

Wireless healthcare, smart port, smart agriculture, smart mining, smart scenic spot, etc.

Figure 1: Differentiated requirements on bandwidths and latencies

The emergence of AI phones, LLMs, and AIGC has spurred the creation of applications that incorporate multi-modal AI-driven interactions, including platforms like Sora and ChatGPT-4o. These applications have transformed the content ecosystem, resulting in increased demands on mobile network uplink and downlink capabilities. For instance, ChatGPT-4o mandates an air interface latency of under 50 milliseconds to deliver an authentic, lifelike experience.

Fortunately, the 5G-A network offers super uplink and downlink speeds and ultra-low latency, making it ideal for emerging services. By leveraging 5G-A's versatile capabilities, operators can take communication experiences to new levels and unlock new opportunities for monetizing differentiated experiences.

2.2 User Foundation for Experience-based Operations: Users Are Pursuing Personalized Experiences and a Sense of Being Prioritized

During the 4G era, network resources were distributed among all users, including both VIPs and common users. Unfortunately, this shared allocation often resulted in resource contention, making it challenging for operators to deliver tailored, high-priority experiences that justified premium pricing. As a result, VIP users often felt dissatisfied with the experiences they paid for, despite their premium status.

The rise of 5G, has ushered in new services with heightened and distinct experience demands. To address these requirements, operators are exploring ways to enhance their 4G network resource allocation methods. Their objective is to efficiently distribute resources while ensuring cost-effectiveness, facilitating a seamless shift from traffic-based to experience-based monetization.

As operators look to monetize differentiated experiences, they must adopt a more sophisticated approach to network resource management. This involves identifying and addressing the unique needs of each user segment, while optimizing resource allocation to ensure maximum efficiency and effectiveness. By doing so, operators can unlock sustainable business growth and realize the full potential of their 5G networks.

A comprehensive survey conducted by Omdia, a renowned global analyst and advisory firm specializing in the tech market, revealed a profound shift in consumer behavior. Over 55% of respondents expressed a willingness to pay a premium for superior service experiences that cater to their individual needs and expectations. Beyond mere satisfaction, VIP users place a high value on being prioritized and recognized as premium customers. To address this need, service providers are exploring innovative strategies to differentiate themselves and deliver exceptional experiences. A notable example of the approaches they adopted is the display of a VIP assurance symbol on terminals during services.

Item	Global	China
Higher downlink rate	55%	55%
Higher uplink rate	40%	51%
Higher connection reliability	39%	47%
Unlimited data	36%	33%
More data	29%	34%
5G-featured applications	20%	22%

Table 1: Consumers' willingness for higher-quality service payment (source: Omdia)

2.3 Network Foundation for Experience-based Operations: Versatile 5G and 5G-A Network Capabilities Empower Leading Operators to Explore Experience-based Service Monetization

The mobile industry has undergone significant transformations since the introduction of 4G, with operators initially capitalizing on user base dividends and later benefiting from traffic dividends. However, the industry is now facing a new reality, with the traffic-based monetization era being challenged by two critical issues: slow global traffic growth and

homogeneous traffic services. According to GSMA's statistics, the global annual traffic growth rate has slowed to below 30%, with only some EM regions in Africa and Latin America exhibiting growth rates over 40%. Concurrently, the proliferation of standardized data services has led to a decline in unit prices per megabyte, even with some developed markets experiencing a decrease of over 50%. In response, operators are compelled to explore innovative strategies for achieving sustainable development and efficient operations. They leverage their network advantages to drive growth and mitigate the impact of these emerging challenges.

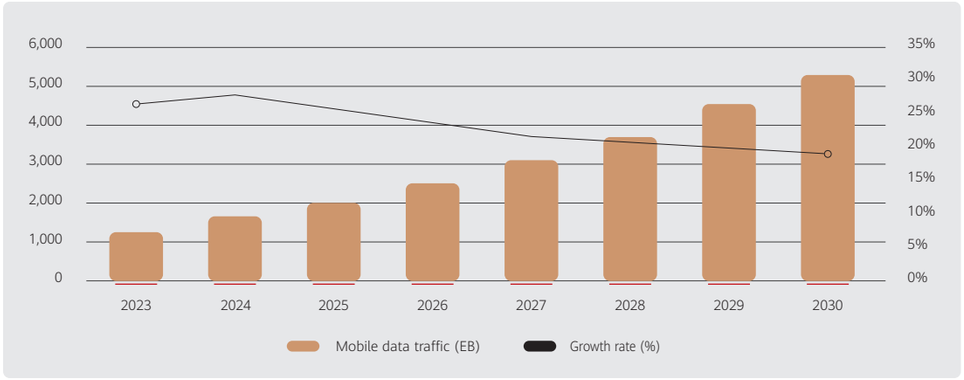


Figure 2: Global mobile data traffic (source: GSMA)

Leading 5G operators in the Asia Pacific region have been at the forefront of innovation, proactively adopting a multi-dimensional experience monetization model to drive revenue growth and improve user experiences. China Unicom's launch of 5G live streaming packages in Guangdong province of China is a notable example, providing super high uplink rates that have attracted over 300,000 users, generating CNY0.3 billion in revenue, and significantly boosting DOU to 6 GB, as well as ARPU by 76%. At the same time, China Mobile has successfully deployed the world's first 5G-A commercial network across Zhejiang province in China, implementing a hierarchical experience assurance mechanism to guarantee seamless experiences for live events, KOL streaming, and mobile office applications, thereby elevating overall communications experiences to a new level.

In an effort to enhance user experiences and drive revenue growth, AIS in Thailand

presented the Living Network hotspot acceleration add-on packages at the end of 2023. This innovative offering provides three distinct acceleration modes: BOOST, which prioritizes higher speeds; GAME, which optimizes for low latency; and LIVE, which delivers super uplink capabilities. Through personalized options that allow users to customize their deterministic experiences, AIS has successfully distinguished its product offerings and elevated the overall service quality for its users.

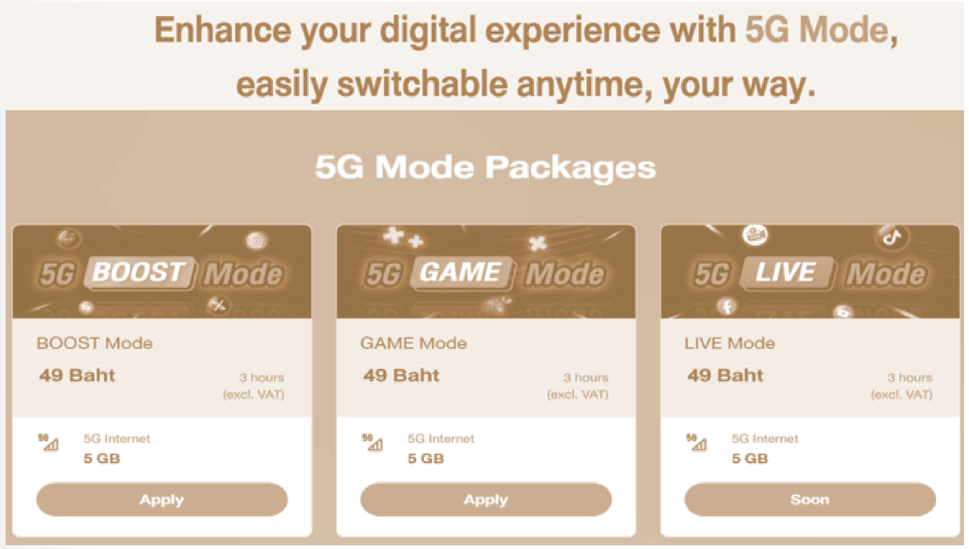


Figure 3: Experience-oriented service packages of AIS in Thailand

Operators in the Middle East are also actively exploring experience-based operations, with leading providers seeking to drive innovation through the hierarchical experience management of FWA users. One such example is du's release of FWA game acceleration packages in the UAE. These packages aim to achieve ultra-low latencies for specified games and also offer prioritization to the members of select OTT apps, including Disney+.

Experience-based operations, which offer a new growth curve for operators, necessitate that their networks provide the following five critical capabilities to ensure that user experiences are optimized:

1. Identifiability: The ability to accurately identify service quality, as reflected by KQIs, high-value users, and service applications in use, enables real-time analysis of user experiences.
2. Reachability: Effective marketing methods, such as targeted short message promotions to high-value users, enable operators to promote add-on subscriptions or package upgrades, increasing revenue opportunities.
3. Guarantee: A tailored indicator system, encompassing specialized indicators for differentiated services and experiences, such as indicators designed for number allocation, allows networks to dynamically guarantee the experiences of various services, ensuring consistent quality of service.
4. Perceivability: Clear communication of assurance effects to users through short messages or symbols displayed on terminals enhances transparency and user satisfaction.
5. Evaluability: A scenario-based fault locating and demarcation system enables networks to rapidly resolve issues using methods such as assurance failure cause analysis and assurance quality analysis, minimizing downtime and maximizing user satisfaction.

3

Trend 2

New Calling Enables Operators to Revitalize Calling Services and Capture Dialer-based Business Opportunities in the Mobile AI Era

To boost innovation in calling services, enhance user experiences, strengthen their user base, and increase revenue, leading operators began exploring service intelligence through New Calling in 2023. They incorporated intelligence and interactivity into existing calling networks to transform their core services and shift towards multi-modal communications. This approach not only offers users a more enjoyable and efficient calling experience but also revitalizes operators' dialer-based service portfolios. As 5G-A and AI technologies unfold in 2024, the calling industry is undergoing a profound transformation, unveiling a vast array of opportunities for operators.

3.1 From Limited Commercial Deployments by Pioneering Operators, New Calling Is Advancing Towards a Global Business Landscape

China Mobile has strategically positioned New Calling as a flagship product and has deployed a nationwide New Calling network serving 50 million users in the second half of 2023. Simultaneously, the company has introduced a suite of innovative services, including intelligent translation, fun calling, visualized voice calling, and AI-powered transcription. Within a remarkably short period of six months, China Mobile has attracted tens of millions of users. This achievement not only enhances communication by making it more efficient and enjoyable but also presents operators with a unique opportunity to acquire new users and generate additional revenue streams. Presently, operators are focused on further developing New Calling into core services, expanding its reach from B2C to B2B markets, and fostering deeper connections with enterprises and third-party players. In May 2024, Huawei and the Jiangsu branch of China Mobile introduced their New Calling-Advanced services, which include enhanced visualized voice calling, intelligent customer service, and real-time voice driven avatar. New Calling-Advanced integrates intelligence at a deeper level, enabling a comprehensive upgrade of B2C and B2B services and providing users with stable, high-definition, interactive, and efficient communication experiences that elevate the calling industry to new heights.

To keep pace with China Mobile's success, other leading operators in China are investing in their own New Calling networks. Globally, operators in Europe, Latin America, the Middle East, and Asia Pacific are experimenting with New Calling and plan to roll it out commercially in 2024. New Calling is becoming the preferred choice for operators seeking to improve user experiences and maximize the potential of their dialer-based service gateways.

3.2 AI Transitions Transparent Information Exchange to Multi-modal Communications, Assisting Operators in Redefining Calling Experiences and Business Models

Since 2024, leading terminal vendors have been actively launching AI-capable phones, aiming to transform the way people communicate, enhance communication efficiency, and meet users' growing demands for AI-powered features. Then, how can operators maximize their 5G-A network potential to stay ahead of the competition in the mobile AI era?

AIGC is integrated into visualized voice calling, enhancing voice calls with multimedia content and expediting content monetization. With visualized voice calling, users can embed customized multimedia elements into voice conversations during the average 90-second call, creating immersive and interactive experiences that captivate audiences, build connections, and drive engagement for individuals and businesses alike. Since its launch, this innovative service has garnered a substantial user base. AIGC transforms content creation by diversifying content formats and boosting production efficiency. Leveraging AIGC, users can create personalized digital characters with diverse styles based on their appearance features and animate these characters using their voices during calls. These digital characters can replicate users' facial expressions in real time, revolutionizing the calling experience with unparalleled levels of engagement, fun, and emotional connection. Enterprise users can also tailor digital characters to embody their brand essence and share targeted information, transforming calls into a new mobile media platform. As New Calling gains momentum, operators can tap into new revenue streams, foster deeper user loyalty, and pivot from voice-only services to content-centric offerings.

Visualized Voice Calling | Real-time Voice Driven Avatar | AI Calling Assistant

Figure 4: Typical New Calling services

AI agents are creatively integrated into calling to develop intelligent assistants capable of understanding communication contexts and supporting decision-making. Leading operators are concentrating on the synergy between AI and calling services. First, AI both enables and enhances intelligent assistants to offer a broad spectrum of services during calls. For instance, these assistants can answer or reject calls on behalf of users, preventing missed calls, harassment, or scams. They can also deliver smart features like translation, reminders, and suggestions to improve communication efficiency and experience. Additionally, they can serve as virtual companions, providing personalized Q&A and emotional support tailored to individual needs. Moreover, AI drives the modernization of enterprise customer services by transforming traditional interactive voice response systems into advanced multi-modal agents. These agents support natural language interactions as well as video and data channels. During customer service interactions, digital humans and customers can communicate naturally, streamlining the service process and improving the efficiency. Before DC-native terminals are widely adopted, multi-modal agents can address key challenges, such as complex DTMF interactions, intricate service navigation, and time-consuming procedures. The convergence of AI and calling services enables operators to develop a unified platform and ecosystem for calling applications, while also boosting user engagement for third-party AI foundation models. This convergence allows calling services to address a wide range of user needs and serves as a key channel for directing user traffic.

3.3 New Calling Standards Have Been Formulated, and the Industry Ecosystem Is Continuously Evolving

In June 2024, 3GPP Release 18 was finalized, standardizing the NG_RTC technology stack. This release defined the architecture, interfaces, and service procedures for IMS DC, and established charging and security mechanisms for real-time communication services, providing a robust technical foundation for the advancement of New Calling.

GSMA and industry partners are collaborating to create a vigorous ecosystem for New Calling. At the end of 2023, GSMA's Board of Directors established the New Calling

Task Force to advance the development of New Calling standards. In the subsequent six months, the Task Force attracted more than 15 members, with the expectation that more industry partners will join and contribute to a vibrant and open industry ecosystem. In July 2024, GSMA officially released the first version of TS.66. It defines the essential APIs that terminals need to support IMS DC, offering a crucial reference for the development of New Calling applications. On October 19, 2023, China Mobile and Huawei announced the establishment of the New Calling Innovation Center. Emphasizing openness, collaboration, and shared success, the center aims to leverage the strengths of telecom operators and the industry to foster new communication technologies, capabilities, and applications. It assists New Calling ecosystem partners in swiftly validating their solutions and launching new services, metamorphosizing calling from voice-only to video- and content-oriented, and ultimately heading towards a thriving industry.

Through the collaborative efforts of stakeholders, major obstacles to the commercialization of DC-native terminals have been largely resolved. In October 2023, the first DC-based call was made using a Huawei Mate 60 Pro over a commercial network. Currently, Chinese operators are conducting joint commissioning tests with terminal vendors and collaborating with leading enterprises to design services. These efforts are effectively integrating DC interaction capabilities into enterprise service workflows, helping enterprises reduce costs and enhance service efficiency. By the end of 2024, the first batch of DC-native terminals is projected to enter commercial use. This will significantly advance the adoption of interactive calling, support operators in establishing a dialer-based application platform, and empower a diverse array of industries.

4

Trend 3

Operators Need to Leverage AI to Enhance O&M Efficiency, Boost Network Reliability, and Address "Historical Debts"

4.1 AI Becomes the Solution to Core Network Efficiency and Reliability Challenges

The continuous evolution of Cloud Native and RATs has resulted in an unprecedented level of complexity in the core network, with significant implications for O&M. For instance, over the past five years, an operator in China has experienced a 1.8-fold increase in the number of NEs, a 75% increase in service types, and a doubling of technology types. This surge has significantly increased the O&M workload, while the number of O&M personnel has remained unchanged, widening the gap between workload and available personnel. Public reports indicate that high network complexity, insufficient O&M manpower, and inadequate O&M methods have compromised core network reliability, leading to frequent accidents and service downtime. Consequently, operators have faced escalating economic and brand losses annually, underscoring the urgent need for innovative solutions to tackle core network complexity and O&M challenges.

Global operators have long pursued cost reduction, efficiency improvement, and network stability. However, the increasing complexity of their networks due to evolving core network technologies and diversifying services has created significant challenges. Despite this growing complexity, the number of maintenance personnel has not kept pace, putting operators under intense pressure due to lagging maintenance resources.

The emergence of Generative AI has injected new momentum into network O&M transformation, making it imperative for operators to fully leverage AI for efficient O&M and high network stability. With the help of foundation models, "digital employees" have been introduced into production systems to replace manual operations. This allows operators to retain existing personnel despite the growth in network quantity and complexity, significantly improving O&M efficiency and addressing "historical debts" accumulated due to insufficient O&M personnel.

Furthermore, the application of twin technologies has made it possible to predict and prevent potential risks, as well as perform simulation-based verification before network changes. This greatly reduces the probability of accidents and ensures high network stability.

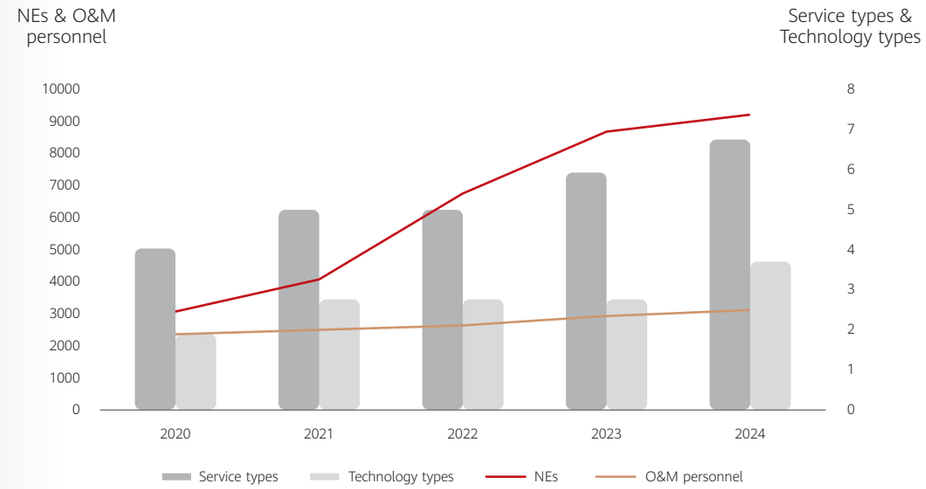


Figure 5: Gap between O&M workload and personnel

The global telecommunications industry is making a concerted effort to leverage intelligence technologies to enhance network efficiency and stability. At the Autonomous Network Industry Summit held during DTW 2024, the TM Forum unveiled the Autonomous Networks Level 4 industry blueprint, urging industry partners to accelerate the adoption of cutting-edge technologies like Generative AI and Predictive AI. This initiative aims to develop systematic automation and intelligence capabilities, propelling operators towards L4 autonomous networks. In response, operators both within and outside China have begun to act. China Mobile has proposed an "AI+" action plan, focusing on E2E automation and harmonizing the integration of foundation and small models with innovative approaches to drive low-cost and efficient operations. China Unicom has launched their Yuanjing LLM, which is being used to develop a range of services for network operations, including knowledge Q&A ChatBot, copilots for task assistance, and scenario-based agents. Meanwhile, multiple tier-1 operators outside China have incorporated AI into their overall strategies and started exploring its potential, with some envisioning the use of Generative AI to transform the telecom industry. For instance, DT believes that Generative AI will bring opportunities to the industry and has proposed a network development strategy centered around enhanced intelligent network autonomy.

4.2 "Digital Employees" Emerged from Foundation Models Alleviate O&M Workforce Gaps

The emergence of Generative AI is driving transformative changes across various sectors, and its impact on core network O&M is particularly significant. Foundation models, which excel in extensive knowledge and capabilities such as natural language understanding, AI emergence, task planning, decision-making, and reasoning, are poised to inject new momentum into core network O&M. However, the highly specialized and scattered knowledge required for this field, encompassing product documents, logs, alarms, KPIs, and signaling, necessitates the development of a foundation model tailored to core network O&M. To address key pain points and improve efficiency, operators must develop copilots and agents to replace manual operations. These should focus on high-value scenarios such as network maintenance, experience assurance, and service enablement, while integrating into critical workflows like monitoring, troubleshooting, complaint handling, and service changes.

The copilots must be designed to play a range of roles, providing natural semantic interaction, intelligent knowledge Q&A, and O&M assistance capabilities that streamline engineer workflows and reduce manual information queries and collaboration. By lowering the operation threshold and improving operation efficiency, copilots can significantly enhance the productivity of each role. Meanwhile, the agents must be able to adapt to various scenarios, understand and break down complex tasks, and invoke the necessary tools and APIs to quickly and independently close tasks, ultimately driving O&M cost reduction and efficiency improvement.

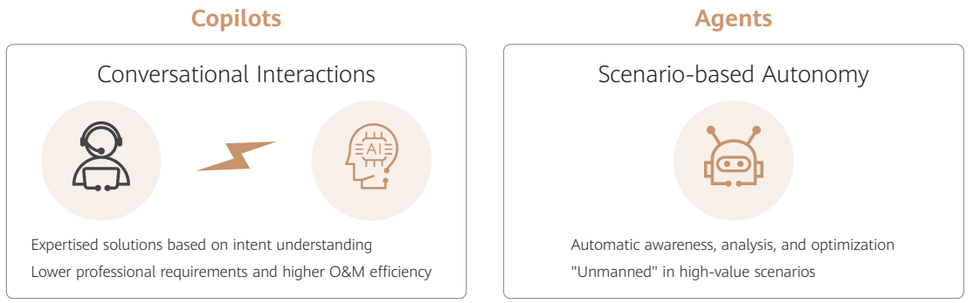


Figure 6: Copilots and agents

4.3 Proactive Risk Mitigation Through Simulation-based Verification Enhances Network Reliability

The core network, located at the heart of the entire network, is a critical component with a far-reaching impact, making its high reliability essential. A thorough analysis of network accidents worldwide reveals that network changes and signaling storms are frequent causes of these incidents. Unfortunately, simulating these high-risk scenarios on a live network is challenging, making it difficult to effectively avoid risks and prevent accidents. However, the emergence of twin technologies has made pre-event simulation and verification possible, enabling operators to evaluate network risks and problems in a controlled environment. For instance, in signaling storm scenarios, fault modes can be injected into a real-time online simulation system based on the twin network, providing valuable insights for capacity expansion and parameter adjustment. Furthermore, intelligent optimization algorithms can automatically generate optimal combinations for network-level flow control parameters.

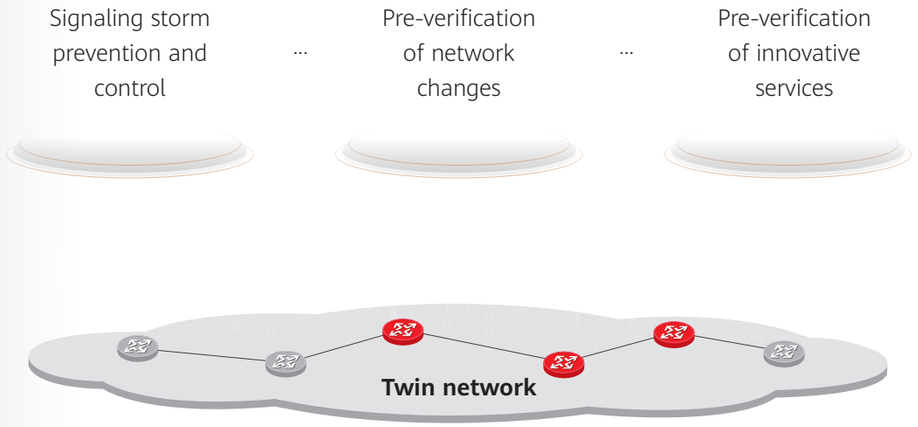


Figure 7: Simulation-based verification

5

Trend 4

5G-A New Capabilities Offer Higher Bandwidth, Deeper Industry Integration, Meshed Networks, and More Connections to Deliver Intelligent, Deterministic Connections

5.1 Higher Bandwidth: HD, Intelligent Services Require Ultra-high Bandwidth

The shift towards HD content in mobile Internet services is fueling the demand for ultra-high bandwidth. According to statistics, the downlink bandwidth for a single cloud gaming flow must be greater than 100 Mbit/s, whereas, the bandwidth for a single XR application flow must be 200 Mbit/s or higher to offer an immersive experience to users. And the bandwidth for a single live broadcasting flow of 4K cameras must be greater than 800 Mbit/s. Speed test software has even higher requirements. For example, Speedtest requires that the single-flow bandwidth be greater than 1 Gbit/s. Additionally, major breakthroughs in terminal connectivity have propelled the development of 3CC-capable terminals to the next level. It is estimated that up to 42 million 3CC-capable terminals will be shipped by the end of 2024, with prices dropping to a range from US$400 to US$600. Furthermore, base station 3CC technology has matured, expanding the air interface channel by 5 to 10 times and achieving 5 to 10 Gbit/s bandwidth for a single flow. The core network currently adopts single-vCPU processing for each flow. However, if the single large-bandwidth flows overwhelm the single-vCPU processing capability, packets will be discarded. To overcome the challenges for the above trend, the core network must support multi-vCPU aggregation, thereby enabling E2E forwarding for ultra-high bandwidth services.

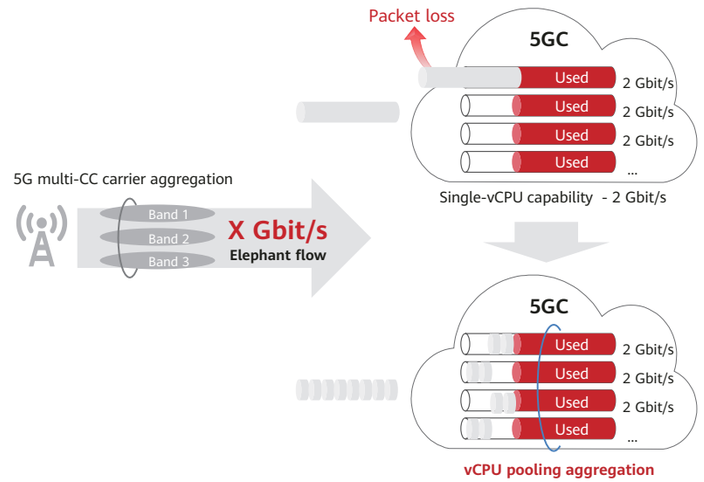


Figure 8: Multi-vCPU aggregation on a core network

5.2 Deeper Industry Integration: 5G and Industrial Internet Penetrate Periphery Assistance and Even Onsite Production to Bring Local-Area Connections to Industrial Sites, Driving New B2B Service Growth

New B2B services grow rapidly to underpin operators' revenue growth. As predicated by GSMA in *Private 5G Industrial Networks 2023*, the revenues of 5GtoB private networks will reach US$109.4 billion in 2030. Pioneering operators like China Mobile are now thriving in this lucrative B2B business cycle, with all their 5G private networks' compound growth rates exceeding 50%, opening up a new industry space.

Leveraging PNI-NPN has become a consensus on network construction. To achieve this, the kite-like solution gives full play to operators' network highlights, such as central network coverage, carrier-grade high reliability, and intensive deployment, monetizing network capabilities. To date, over 15,000 public networks have been deployed worldwide for private use, spanning multiple industries, such as education, healthcare, government, and energy. For example, the State Grid Corporation of China has deployed a grid-dedicated slice and an independent core network on the existing operator network. This separates the management domain from the service domain and automates precise load control as well as the control on distributed photovoltaic modules, improving efficiency by 50%.

5G and Industrial Internet bring local-area connections to industrial sites. There are over 5000 5G private networks for industrial manufacturing worldwide, bringing campus assistance connections to industrial sites, from industrial extranets to industrial intranets. In addition, certain functions are introduced. For example, the dual-system hot backup function and the dual fed and selective receiving function are introduced to deliver industrial-grade high reliability; 5G LAN-based full interconnection (featuring wired-to-wireless transformation without changing original applications, topology, configurations, and performance) and 5G indoor high-precision positioning are provided for industrial workshop devices to step up flexible manufacturing. Furthermore, deterministic experience assurance is provided for OT devices while 5G penetrates the core production system, bolstering smart manufacturing through wireless networks in more scenarios. In 2023, China Unicom and Huawei built the world's first ultra-reliability, ultra-low latency, and flexible production line for Great Wall Motor. This production line provides a 5G-based deterministic low latency for robotic arms in automobile manufacturing, significantly reducing the downtime from 15 days to just 2 days.

5.3 Meshed Networks: Cross-Area Connections Shift from Connected People to Connected Things, Connected Vehicles, and Connected Homes, Evolving Towards MEC-powered Fully Meshed Connections

The evolution from connected people to connected things and then to connected vehicles and connected homes helps construct human-centered, vehicle-centered, and home-centered communications networks, respectively. The proliferation of personal terminals (PCs, mobile phones, and wearable devices) and IoT devices has triggered a significant surge in network connections, fostering a highly interconnected environment that spans humans, vehicles, and homes, as well as their interactions. To satisfy these requirements, the network needs to evolve from B2C-only connections to fully meshed connections between humans, vehicles, homes, and enterprises. In the future, it also needs to manage thousands of large-, medium-, and small-sized connections and fulfil the interworking requirements between various networks.

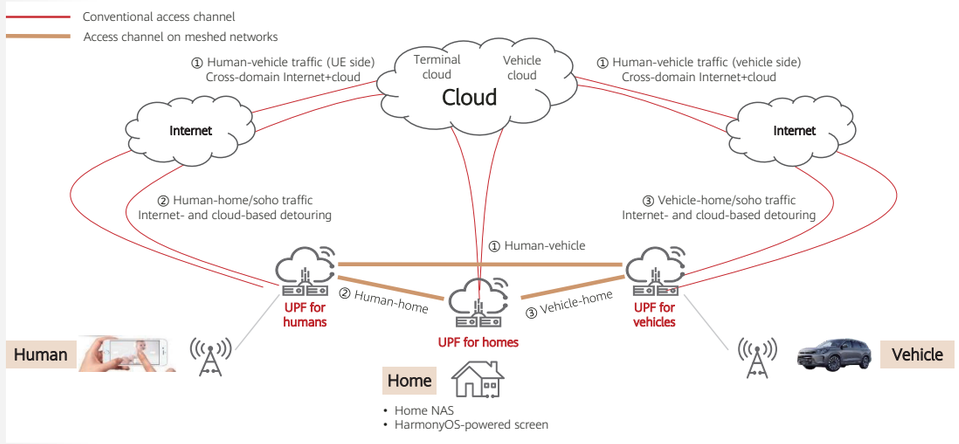


Figure 9: One-hop access to home/vehicle services and conventional multi-hop access to home/vehicle services

Human-home connections enable people to remotely access smart home or enterprise office devices as smart terminals develop. Human-enterprise connections provide easy access to ubiquitous mobile office networks of campuses and enterprises. Human-vehicle connections unlock remote access to the vehicle-centric third space. Vehicle-home connections and mutual control between vehicles and homes enable people to remotely control home terminals through the control screens inside vehicles and also remotely control the climate control system and ventilation of vehicles through the control screens at homes. Furthermore, they facilitate information sharing between vehicles and between homes.

Only meshed networks can help meet these various interconnection requirements.

5.4 More Connections: 5G-A Enables a Single Network to Achieve Multiple Capabilities, with RedCap and A-IoT Charting a New Chapter for B2B Services

As IoH connections evolve to IoT connections, 5G-A is continuously meeting the diversified requirements of hundreds of billions of IoT connections. In the 4G era, NB-IoT technology enables low-power-consumption devices to interconnect with each other on wide-area networks. This technology offers wide coverage, large capacity, low power consumption, and cost-effectiveness, making it ideal for scaling up in various vertical industries that require up to 100 kbit/s bandwidth, such as remote meter reading and smart street lamps. LTE-M is designed for smart wearables and industrial sensing and requires a maximum of 1 Mbit/s bandwidth. It occupies more bandwidth than NB-IoT and therefore consumes more power. In the 5G/5G-A era, the network offers higher connection speeds than those in the 4G area, and IoT connections evolve accordingly. On the one hand, RedCap enables a network access rate of 10 Mbit/s to 100 Mbit/s and applies to IoT scenarios that are relatively sensitive to bandwidth, such as video surveillance and smart watches. On the other hand, A-IoT is introduced to provide lightweight, low-speed, and low-cost IoT access. This technology applies to IoT scenarios that are less sensitive to bandwidth, such as warehouse counting, logistics monitoring, as well as temperature, humidity, and pressure monitoring. Below are the details of these technologies.

RedCap is the optimal technology for medium- and high-speed IoT scenarios, empowering network evolution from IoH to ubiquitous, all-scenario connections. The RedCap industry ecosystem can be scaled up for commercial use and has entered the large-scale development stage. RedCap is benchmarked against 5G eMBB and features lower costs (similar to those of 4G for the same scale) and lower power consumption (10% to 20% lower in connected mode). As such, RedCap applies to multiple scenarios, such as smart power, video surveillance, smart manufacturing, IoV, and wearable devices. Increasing connections will drive a steady decrease in module prices and the continuous enrichment of chips and modules. By 2024, over 100 industry terminals will be launched. The three tier-1 operators in China will commercially use RedCap on a large scale in 2024. Leading operators outside China will also start the commercial deployment of RedCap.

A-IoT supports a range of typical application scenarios, including smart warehousing, smart manufacturing, and logistics tracking. It also supports three types of A-IoT tags: A (passive), B (semi-passive), and C (active). Currently, the prototype terminal samples for this technology have been tested and are ready for use, scenario-based applications have entered the cultivation period, but the terminals for this technology are still developing slowly. Therefore, it still takes time for commercial use, mass production, and full maturity of this technology.

To achieve ultra-low power consumption of A-IoT, the core network leverages the lightweight NAS communication protocol to communicate with A-IoT tags and identify terminal types according to tags A, B, and C. A-IoT-based access capabilities can be flexibly deployed for different types of networking in indoor, outdoor, local areas, and wide areas, whereby it can manage and control a vast array of terminals. This fully unleashes the potential of 100 billion A-IoT connections.

A single network with multiple capabilities is a game-changer for operators, enabling them to improve efficiency, increase revenues, and maximize value. Across thousands of B2B industries, operators must harness the full potential of 5G technologies to provide converged access, such as 4G, NB, 5G, RedCap, and A-IoT, through a single network, facilitating network scalability and evolution.

6

Trend 5

5G-A Goes Beyond Connections; It Integrates Communications, Computing, and Storage, Facilitating Content Storage on Operator Networks and Extending into Low-Altitude Economy

Propelled by advancements in technology and business innovation, substantial progress has been achieved in glasses-free 3D, AR, MR, and VR. The 5G-A core network converges communications, computing, and storage, paving the way for operators and partners to explore the low-altitude economy and apply it to more scenarios like intelligent transportation.

At the global launch of the world's first 5G-A commercial deployment in March 2024, China Mobile, Huawei, Coolpad, iFLYTEK, Cocos, NOLO, and other partners established the 5G-A Glasses-Free 3D Industry Alliance. They have introduced a wide range of glasses-free 3D products to the market, including smartphones, tablets, in-car displays, laptops, and monitors. Their achievements span content applications, technology platforms, and ecosystem capabilities. Behind these achievements lies a robust network. China Mobile's 5G-A commercial deployment provides strong network support for glasses-free 3D services. With high-speed connectivity, extensive coverage, low latency, and AI technologies, users can expect more immersive and lifelike experiences.

The XR market is thriving, with leading technology companies consistently introducing VR/AR smart glasses and headsets that offer immersive user experiences through spatial computing technologies. For example, Apple has launched Apple Vision Pro, which is equipped with Apple's M2 and new R1 processors, dual 4K micro-OLED displays, and a specially designed catadioptric lens system, providing users with crystal clear and lifelike visual effects. Moreover, it seamlessly tracks eye and hand gestures, enabling natural interactions between the physical and virtual worlds. This product is warmly embraced by enterprises, with over 50 percent of Fortune 100 companies purchasing the units. It is extending footprints across diverse sectors, including healthcare, manufacturing, education, and automotive.

The integration of large models and spatial computing is indeed a game-changer in merging the physical and digital worlds. With its high uplink and downlink broadband, low latency, and HCS capabilities, 5G-A will establish a robust foundation for large models and spatial computing.

6.1 Short Videos Go HD; Operators and OTT Businesses Promote Content Storage on Operator Networks

Short videos have emerged as a dominant service on the mobile Internet. In 2023, the DAU of short video platforms exceeded 2.6 billion, with Chinese users accounting for approximately 90%. This underscores the pivotal role that short videos play in our daily lives.

Empowered by 5G, short videos are increasingly likely to be available in HD. Analysis reveals that within a 5G network, the top 18% of popular short videos — accounting for approximately 90% of the total view counts — are H.265 encoded. The surging popularity of HD content places greater demands on bandwidth. For instance, to serve 3 million Douyin users with HD content, the short video giant needs a CDN bandwidth of about 60 Gbit/s. And as its user base continues to grow, the need for even higher bandwidth becomes paramount.

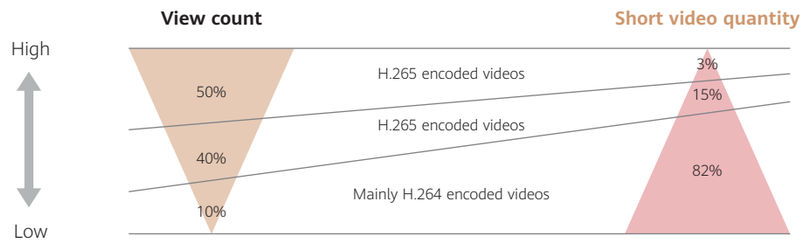


Figure 10: H.265 content proportion (source: CSDN)

Beyond network infrastructure, HD content also poses challenges for OTT companies in terms of network costs. A mere 10% increase in the proportion of 1080p videos leads to a 16% rise in daily wireless traffic, resulting in significant CDN bandwidth costs. Consequently, OTT platforms may opt to lower video bitrates to mitigate bandwidth expenses. However, this decision may adversely affect user experiences, leave the demand for HD content unmet, and ultimately impact the development of the OTT platform.

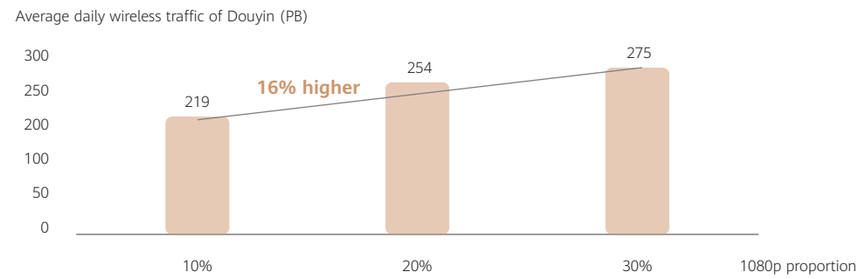


Figure 11: Influence of 1080p video proportion of Douyin on data traffic

The bandwidth issue may also impact telecom operators. If OTT platforms lower video bitrates, users may be not inclined to upgrade their traffic service packages, and the DOU increase may stall and hit a plateau. As users consume less data with lower-bitrate videos, the overall growth in data traffic consumption slows, subsequently impacting operators' data revenue and the sales of premium traffic service packages.

To address the aforementioned challenges, operators and OTT platforms are exploring media relay solutions to achieve mutual benefits for users, OTT platforms, and operators. By integrating media relay technologies into their networks, which can relay, cache, and schedule media content encapsulated using HTTP 1.1/2.0/3 protocols, operators can significantly enhance video stream transmission efficiency and quality. With the media relay capabilities on operators' networks, OTT platforms can transmit video content more efficiently and rapidly to users, with lower latency and packet loss, thereby improving users' viewing experiences.

The deployment of media relay solutions on operators' networks offers several benefits. By doing so, it reduces redundant bandwidth consumption while maintaining video quality, effectively alleviating CDN bandwidth pressure. In this way, OTT platforms no longer need to lower bitrates to conserve bandwidth, avoiding negative impacts on user experiences. Importantly, this approach does not affect user data traffic consumption. As user data traffic continues to rise, operators' DOU also grows, leading to increasing revenue and encouraging traffic plan upgrades. Additionally, deploying media relay on operators' UPFs has inherent advantages over deploying it on the CDN outside

operators' networks. Firstly, operators can coordinate the forwarding and scheduling capabilities of UPFs and RAN for various scenarios, such as frame-level aggregation and forwarding and priority-based scheduling of frame packets, to reduce service latency. Secondly, this helps operators further collaborate networks and services. UPF-based media relay can utilize the intelligent capabilities of the network to provide each user with content resolution that best matches their network conditions, achieving more accurate service experience awareness. Furthermore, the intelligent network dynamically adjusts bitrates through media relay based on network changes, ensuring smooth experiences without stuttering.

6.2 HCS, Practiced in Various Scenarios, Moves Toward Standardization and Industrialization

HCS technologies help realize one network for multiple purposes. With its multifaceted capabilities, it is being deployed in various scenarios such as low-altitude, maritime, and road transportation, serving as a significant growth engine for operators. The low-altitude economy, experiencing rapid growth, is emerging as a crucial driver of economic growth, while also demanding advanced sensing and operational capabilities.

The low-altitude economy has vast market potential and diverse application scenarios. It covers various low-altitude operations such as logistics distribution, tourism, urban management and inspections, agricultural and forestry protection, emergency rescue, and much more. As of 2023, the market scale of China's low-altitude economy had surpassed CNY500 billion, and is projected to reach CNY2 trillion by 2030. According to the *China Low-Altitude Economy Development Report* released by CCID Consulting, the low-altitude economy has immense three-tier trillion CNY development potential, including low-altitude aircraft manufacturing, general airport construction, and low-altitude new infrastructure. Among these, low-altitude infrastructure is fundamental to the growth of the entire sector, which means that a low-altitude Internet of intelligence needs to be built for communications and management of low-altitude aircrafts.

To date, the low-altitude economy has made remarkable strides across various dimensions.

In terms of policy support, China incorporated the low-altitude economy into the *National Comprehensive Three-dimensional Transportation Network Planning Outline* in February 2021, and started to enact the *Interim Regulation on the Administration of the Flight of Unmanned Aircraft* from January 1, 2024. In March 2024, China's MIIT, MOST, MOF, and CAAC together issued the *Implementation Plan for Innovation and Application of General Aviation Equipment (2024-2030)*, aiming at fostering a new growth engine for the low-altitude economy.

In terms of technology readiness, 5G-A networks with large bandwidth, low latency, and extensive connectivity capabilities are pivotal in supporting the development of the low-altitude economy. The 5G-A network is well suited for the real-time control and data transmission of drones during flight, facilitating their use in logistics, inspection, geographical surveying and mapping, aerial photography and live streaming, and other application scenarios. It ensures the rapid and stable transmission of flight data, images, and video content. Furthermore, 5G-A introduces HCS capabilities into mobile networks, enabling detection, tracking, and sensing on the network. As a critical part of low-altitude infrastructure, the 5G-A network not only provides communications services but also performs positioning, ranging, and speed measurement for drones, thereby supporting the monitoring and management of low-altitude aircrafts.

In terms of standards and protocols, 3GPP SA WG1 has finalized the requirements study for HCS scenarios in Release 19. SA WG1 initiated discussions on the HCS network architecture project during Release 18, while RAN WG1 commenced research on HCS channel models in Release 19. On March 19, 2024, IMT2020 published the *5G-Advanced Integrated Sensing and Communication Network Architecture Research Report (Second Edition)*, which systematically addresses 14 key issues, pushing HCS towards standardization and industrialization. This report not only outlines the comprehensive network architecture for HCS services, defining the capabilities and reporting requirements for the SF on the core network and sensing-capable devices, but also establishes explicit guidelines for the selection of sensing nodes, data transmission, sensing methods, and wide-area service continuity. Furthermore, it explores service policy provisioning, billing, and security privacy concerns pertinent to operators,

particularly investigating the monetization method of charging application users when device-based billing is impractical.

In terms of scenarios and application cases, HCS is being extensively practiced and innovated within the low-altitude economy and low-altitude security sectors. In the low-altitude economy sector, the feasibility of low-altitude drone delivery has been demonstrated by over 30 projects, including SF Phoenix Wings in Guangdong, Meituan Delivery in Guangdong, and the Shanghai-Zhoushan cross-sea route. Without ground traffic constraints, delivery times are halved compared with land delivery, effectively doubling delivery efficiency. In the low-altitude security sector, more than 20 projects, such as Hangzhou Olympic Sports Center, Shenzhen Dayun Arena, and Fujian low-altitude security, have shown that 5G-A can help detect aircrafts and identify unauthorized flights with a detection rate exceeding 95%, reducing monitoring costs by about 40%. To expand the application scenarios of HCS, the sensing capabilities of various devices, including cameras, radars, and base stations, also need to be integrated into the core network, which will help enhance sensing accuracy and extend into more scenarios such as IoV and factory logistics.

As a pivotal direction for the development of communications technologies, HCS will profoundly influence the evolution of the low-altitude economy, smart transportation, and intelligent manufacturing.

7

Trend 6

A Decade of NFV Exploration Leads Operators to Embrace Telco Cloud as a Strategic Imperative to Realize E2E Automation

As 5G-A and AI technologies rapidly mature, telco clouds are undergoing a profound transformation. Operators are accelerating their shift from virtualization to cloud-native, transitioning from VNFs to CNFs. Moreover, the rise of innovative services demands diverse computing power, positioning new computing architectures, such as XPUs, at the forefront. Additionally, the introduction of new services has further complicated the networks, creating an urgent need for automation within telco clouds. Over the past decade, operators have achieved full cloudification in accordance with ETSI standards. And when looking ahead, the industry will continue to evolve and innovate under unified standards and the ongoing advancements in containerization and AI.

7.1 Converged Architecture and Unified ICT Infrastructure

Driven by the growing demand for advanced technologies, high resource utilization, and smooth network evolution, operators are transitioning from virtualization to containerization, a pivotal direction in the development of the telco cloud infrastructure. The shift from VNFs to CNFs not only enhances application isolation and elastic scalability but also significantly improves resource utilization and deployment efficiency. With the maturation of ETSI's container standards, some operators are exploring the use of bare metal containers in new projects. However, due to the widespread adoption of VMs in the telecom cloud market (as shown in Figure 12) and specific security and storage requirements for certain telecom services, VMs and containers will continue to coexist in telecom networks for the foreseeable future.

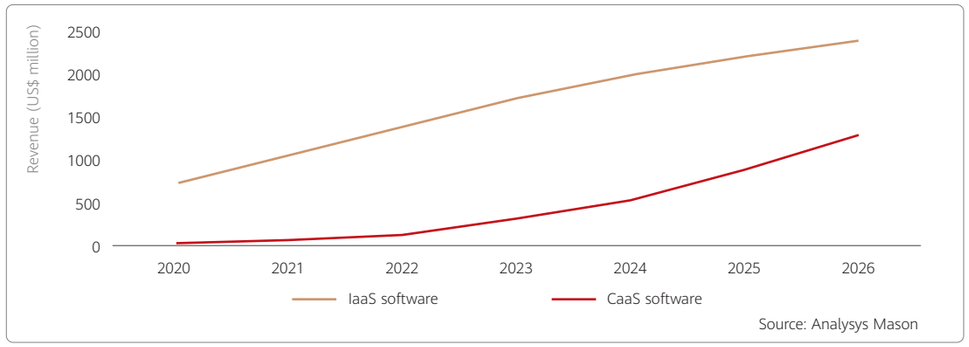


Figure 12: IaaS/CaaS product revenue estimation

Against this backdrop, the dual-engine architecture emerges as a crucial innovation, enabling the seamless introduction of bare metal containers into live networks. By facilitating unified scheduling across multiple types of computing resources — including VMs, containers, and bare metal servers — this architecture effectively fulfills diverse service requirements. According to a report from ABI Research, a typical 5GC bare metal container project employing the dual-engine architecture can achieve a CAPEX reduction of approximately 32% compared to the conventional approach (through new deployment), thereby safeguarding operators' investments on the live network.

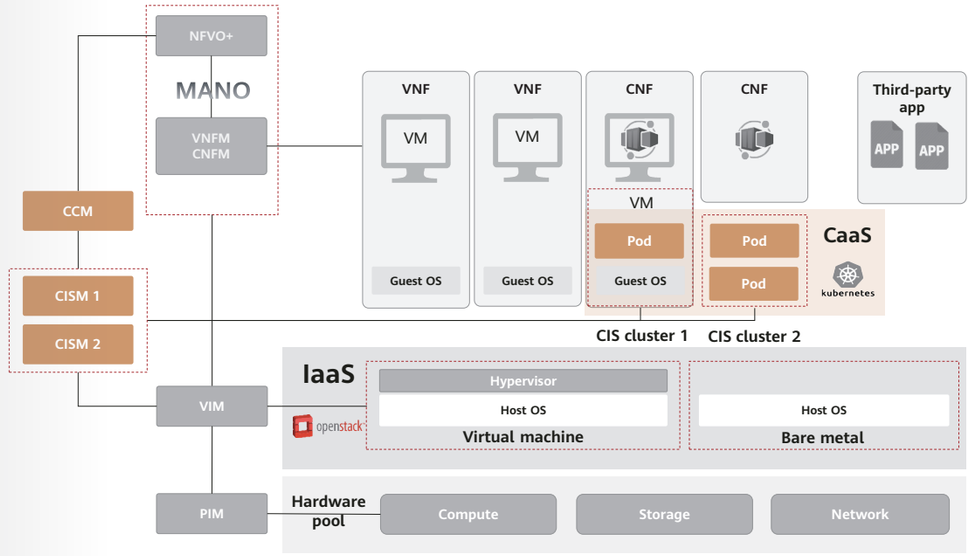


Figure 13: Dual-engine architecture

In today's rapidly evolving telecom landscape, diversified computing power has become the cornerstone for innovation. Typically, CPUs handle a wide range of general-purpose computing tasks, while dedicated processors (such as GPUs, NPUs, DPUs, and FPGAs) are optimized for specific high-performance tasks (such as graphics and video processing and machine learning) to improve efficiency and performance. As new application scenarios emerge and intelligence is integrated into telecom networks, these networks must process diverse types of workloads, leading to increased network complexity. At the Computility Network Conference 2024, China Mobile unveiled the cutting-edge DPU chip (named Da Yun Pan Shi), featuring an impressive bandwidth of 400 Gbit/s. This DPU will

be extensively used in China Mobile's data center construction to support both general-purpose computing and intelligent computing. Furthermore, mainstream operators are progressively introducing heterogeneous computing power to adapt to the high-performance computing requirements of the diversifying services, and provide more secure, reliable, and efficient support for high-parallel computing, edge computing, big data processing, and AI foundation model training. In the future, a diversified computing architecture will be the foundation for telecom service innovation.

7.2 Unlocking E2E Automation and Unleashing the Full Potential of Elasticity and Agility

According to TM Forum's research, 91% of global operators have identified network automation as a strategic priority. After a decade of development, operators have reached a consensus that the technical architecture, whether through hierarchical decoupling or public cloud, is only a pathway towards cloudification. The essence of a telco cloud lies in E2E automation, which aims to minimize network faults caused by manual errors and facilitate rapid fault locating and recovery of services, thereby achieving agile delivery and efficient O&M.

Let's delve into the two focuses of a telecom network's E2E automation.

1. Automatic network changes: With the advent of new services and more frequent network changes in the 5G-A era, operators need the entire process — from requirement identification to design, testing, deployment, rollout, and capacity expansion — to be visible, controllable, and manageable. This ensures secure network changes, prevents human errors, and enables agile delivery.

2. Cross-layer O&M automation: The layered architecture introduced by cloudification complicates routine O&M. A preferred telco cloud should enable operators to visualize network topologies, including computing, storage, and network hardware resources, the cloud platform infrastructure, VNFs/CNFs, and the network management layer. It should also allow for centralized resource management and visualization of the impact caused by cross-layer faults, assisting operators in quickly locating and rectifying the faults. In this

context, a full-stack architecture with a global view facilitates cross-layer collaboration and is the optimal choice for E2E automation.

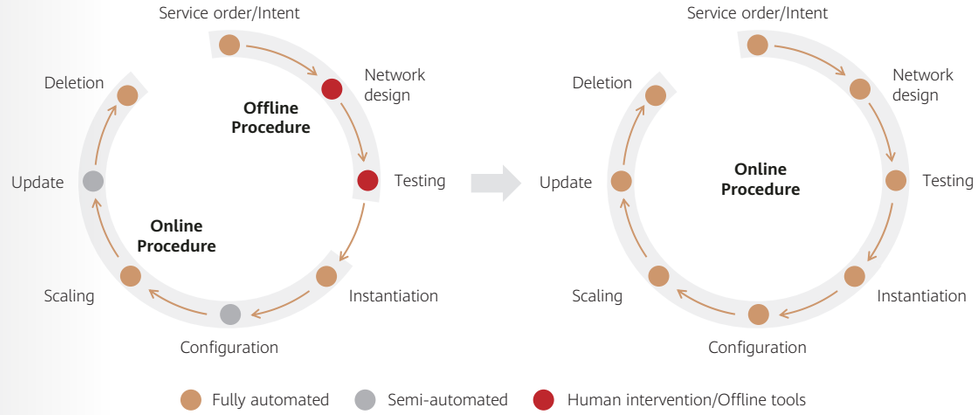


Figure 14: E2E automation process (source: ETSI's 10th anniversary whitepaper)

As we stride towards the era of intelligent O&M, operators should leverage AI to further enhance fault detection, root cause analysis, and self-healing capabilities. This will lead to greater energy savings, improved service quality, and an enhanced user experience.

7.3 Aligning with a Unified Standard and Collaborating on Best Practices

Over the past decade, the large-scale commercial deployment of cloudified core networks has been made possible under the guidance of the ETSI ISG NFV. By incorporating IT-originated open-source technologies, such as OpenStack and Kubernetes, ETSI has proposed a rock-solid network framework for the telco cloud platform. Moreover, it has made multiple enhancements to ensure that telco clouds can deliver carrier-grade reliability and performance. For example, the CPU core pinning technology is utilized to guarantee carrier-grade low latency, and the hitless upgrade

capability is introduced to reduce risks associated with network changes and ensure uninterrupted services during the cloud platform upgrades.

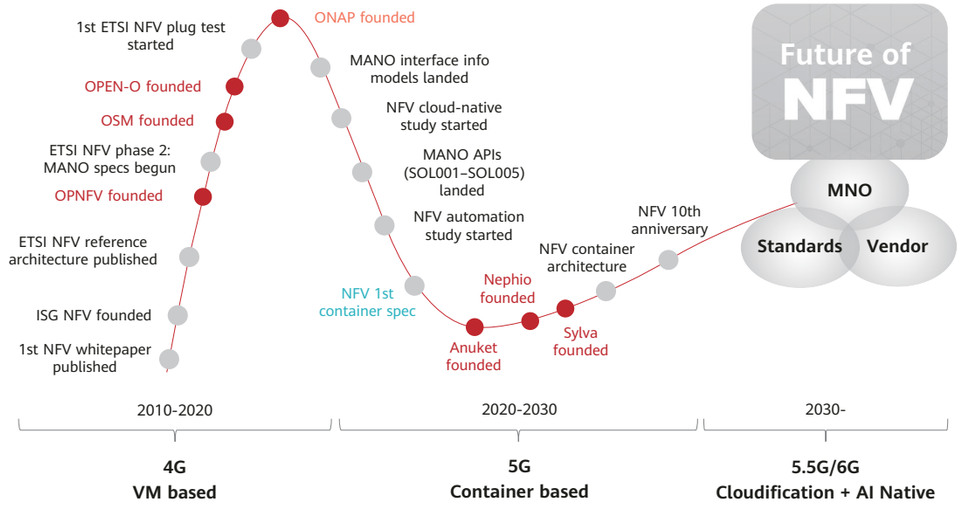


Figure 15: Continuous evolution of NFV standards

ETSI envisions a future where telco clouds evolve from container cloud-native to AI-native platforms. It has initiated standards research projects aimed at developing new architectures, managing intelligent computing resource pools, exploring lightweight virtualization technologies, and advancing intelligent O&M practices. Continued development of NFV-rooted telco cloud standards will undoubtedly help operators, telecom equipment vendors, and IT vendors reach a consensus on a unified functional framework and interaction model, boost more platform and O&M innovations, and foster best practices throughout the industry.

8 Summary and Prospect

As we look to the next decade, the mobile AI era is poised to transform the world in profound and far-reaching ways. With 5G-A as a solid foundation, we will continue to push the boundaries of what is possible, leveraging the convergence of AI and 5G-A to create a truly intelligent world. The combination of 5G-A and AI foundation models will bring about disruptive changes to devices, human-device interactions, business models, and digital service entry points, accelerating industry digital transformation and permeating intelligence into every aspect of life.

This white paper has introduced the key trends brought about by AI and the 5G-A core network in the mobile AI era, and has provided insights into how operators can leverage these trends to develop and monetize the 5G-A core network. By redefining business models and service experiences through the convergence of AI with devices, networks, and services, operators can unlock new revenue streams and create value for their customers.

The 5G-A core network, featuring service intelligence, network intelligence, and O&M intelligence, will help operators fast-track their digital transformation and enhance business success. By harnessing service intelligence, operators can reimagine service entry points, while network intelligence enables experience-driven operations and innovative business models. Meanwhile, O&M intelligence revolutionizes the O&M framework, allowing digital employees and experts an unparalleled position to improve both network reliability and efficiency.

The 5G-A core network promises to elevate connectivity with significantly increased bandwidth, profound industry integration, comprehensively meshed networks, and numerous connections. Moreover, this cutting-edge infrastructure integrates advanced HCS, computing, and storage capabilities, allowing operators to pioneer novel applications, particularly within the scope of the low-altitude economy, and unlock untapped revenue streams in the mobile AI era.

We believe that the mobile AI era presents a tremendous opportunity for collaboration and innovation among industry partners. We call on all stakeholders to join us in working towards a shared vision of a truly intelligent world, where consumers enjoy brand-new communication experiences, vertical industries benefit from highly reliable and intelligent connections, and operators seize the opportunities to unlock new business value. Together, we can build a brighter future for all.

Acronyms and Abbreviations

Acronym and Abbreviation	Full Spelling
3CC	3 Component Carriers
3D	Three-Dimensional
3GPP	3rd Generation Partnership Project
5G-A	5G-Advanced
5GC	5G Core
AI	Artificial Intelligence
AIGC	AI Generated Content
A-IoT	Ambient IoT
API	Application Programming Interface
AR	Augmented Reality
ARPU	Average Revenue Per User
B2B	Business to Business
B2C	Business to Consumer
CAAC	Civil Aviation Administration of China
CaaS	Containers as a Service
CAPEX	Capital Expenditure
CCID	China Center for Information Industry Development
CCM	CIS Cluster Management
CDN	Content Delivery Network
CISM	Container Infrastructure Service Management
CNF	Cloud Native Network Function
CNFM	Cloud Native Network Function Manager

CPU	Central Processing Unit
DAU	Daily Active Users
DC	Data Channel
DOU	Dataflow of Usage
DPU	Data Processing Unit
DTMF	Dual-Tone Multi-Frequency
DTW	Digital Transformation World
E2E	End-to-End
EM	Emerging Market
eMBB	Enhanced Mobile Broadband
ETSI	European Telecommunications Standards Institute
FPGA	Field Programmable Gate Array
FWA	Fixed Wireless Access
GPU	Graphics Processing Unit
GSMA	Global System for Mobile Communications Association
HCS	Harmonized Communication and Sensing
HD	High Definition
HTTP	Hypertext Transfer Protocol
IMS	IP Multimedia Subsystem
IMT2020	International Mobile Telecommunications-2020
IoH	Internet of Humans
IoT	Internet of Things
IoV	Internet of Vehicles
ISAC	Integrated Sensing and Communication
ISG	Industry Specification Group
KOL	Key Opinion Leader
KPI	Key Performance Indicator

KQI	Key Quality Indicator
L4	Level-4
LAN	Local Area Network
LLM	Large Language Model
LTE-M	LTE for M2M
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MIIT	Ministry of Industry and Information Technology
MNO	Mobile Network Operator
MOF	Ministry of Finance
MOST	Ministry of Science and Technology
MR	Mixed Reality
NAS	Non-access Stratum
NB	Narrowband
NB-IoT	Narrowband Internet of Things
NE	Network Element
NFV	Network Function Virtualization
NG_RTC	Next Generation Real Time Communication
NPU	Network Process Unit
O&M	Operation & Maintenance
OLED	Organic Light-Emitting Diode
ONAP	Open Networking Automation Platform
OPEN-O	OPEN-Orchestrator Project
OPNFV	Open Platform for NFV
OSM	Open Source MANO
OT	Operation Technology

OTT	Over the Top
PB	Petabyte
PC	Personal Computer
PIM	Physical Infrastructure Manager
PNI-NPN	Public Network Integrated Non-public Network
Q&A	Question and Answer
RAN	Radio Access Network
RAT	Radio Access Technology
RedCap	Reduced Capability
SA	Standalone
SF	Sensing Function
TS	Technical Specification
UAE	United Arab Emirates
UAV	Unmanned Aerial Vehicle
UPF	User Plane Function
vCPU	Virtualized CPU
VIM	Virtualized Infrastructure Manager
VM	Virtual Machine
VNF	Virtual Network Function
VNFM	Virtual Network Function Manager
VoLTE	Voice over LTE
VR	Virtual Reality
WG	Work Group
XR	Extended Reality